NATIONAL RESEARCH UNIVERSITY

Лекции по эконометрике №1, 2, 3 модуль

Метод максимального правдоподобия

Демидова

Ольга Анатольевна

https://www.hse.ru/staff/demidova_olga

E-mail:demidova@hse.ru

11.01.2021, 18.01.2021,

План лекции

1) Метод максимального правдоподобия
2) Примеры применения метода максимального правдоподобия
3) Свойства метода максимального правдоподобия
4) Проверка гипотез с помощью теста Вальда, теста отношения правдоподобия, теста множителей Лагранжа
5) Выбор моделей с помощью критериев AIC, BIC

Метод максимального правдоподобия

Основное предположение:

Известен закон распределения случайных величин, зависящий от набора параметров.

Оценки этих параметров подбираются таким образом, чтобы вероятность получить имеющийся набор данных была максимальной.

$$L(\theta) = L(x_1,\ldots,x_n,\theta) = P(X_1 = x_1,\ldots,X_n = x_n \mid \theta) \to \max$$

Пример 1. Найти оценку максимального правдоподобия параметра пуассоновского распределения по выборке

Решение:

$$L(\lambda)=L(x_1,\ldots,x_n\,|\,\lambda)=P(X_1=x_1\,|\,\lambda)\cdot\ldots\cdot P(X_n=x_n\,|\,\lambda)=$$

$$=\frac{e^{-\lambda}\cdot\lambda^{x_1}}{x_1!}\cdot\ldots\frac{e^{-\lambda}\cdot\lambda^{x_n}}{x_n!}=\frac{e^{-n\lambda}\cdot\lambda^{\sum\limits_{i=1}^{n}x_i}}{x_1!\ldots x_n!}$$

$$l(\lambda)=\ln L(\lambda)=-n\lambda+\sum_{i=1}^{n}x_i\cdot\ln\lambda-\ln(x_1!\ldots x_n!)$$

Метод максимального правдоподобия

Необходимое условие первого порядка:

$$l'_{\lambda}(\lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$$

Достаточное условие:

$$l''_{\lambda}(\lambda) = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$$

$$\hat{\lambda}_{\mathrm{ML}} = \overline{X}$$

Пример 2. Найти оценку максимального правдоподобия математического ожидания μ и дисперсии σ^2 по выборке $x_1,\ldots,x_n$

Решение:

$$L(\mu,\sigma^2)=L(x_1,\ldots,x_n\,|\,\mu,\sigma^2)=f_{X_1,\ldots,X_n}(x_1,\ldots,x_n\,|\,\mu,\sigma^2)=f_{X_1}(x_1\,|\,\mu,\sigma^2)\cdot\ldots\cdot f_{X_n}(x_n\,|\,\mu,\sigma^2)=$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x_1-\mu)^2}{2\sigma^2}\right\}\cdot\ldots\cdot\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x_n-\mu)^2}{2\sigma^2}\right\}=$$

$$=\frac{1}{(2\pi)^{n/2}\left(\sigma^2\right)^{n/2}}\cdot\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$l(\mu,\sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Метод максимального правдоподобия

Пример 2. Необходимое условие первого порядка:

$$\frac{\partial l(\mu,\sigma^2)}{\partial \mu} = \frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \quad \Rightarrow \quad \hat{\mu} = \bar{X}$$

$$\frac{\partial l(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \qquad \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow \hat{\mu}_{ML} = \bar{X}, \quad \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{X})^2$$

Достаточные условия:

$$H(\hat{\mu}_{ML},\hat{\sigma}^2_{ML}) = \begin{pmatrix} -\dfrac{n}{\hat{\sigma}^2_{ML}} & 0 \\ 0 & -\dfrac{n}{2(\hat{\sigma}^2_{ML})^2} \end{pmatrix}$$

Метод максимального правдоподобия для регрессионных моделей

$$Y = X\beta + \varepsilon$$

$$\varepsilon \sim N(0_n, \sigma^2 I_n)$$

$$L(\ \beta,\sigma^2) = f(\varepsilon \mid \beta,\sigma^2) = \frac{1}{(2\pi)^{n/2}\left(\sigma^2\right)^{n/2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right\} =$$

$$= \frac{1}{(2\pi)^{n/2}\left(\sigma^2\right)^{n/2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}$$

$$l(\ \beta,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)$$

$$l(\ \beta,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(Y'Y - 2\beta'X'Y + \beta'X'X\beta)$$

Метод максимального правдоподобия для регрессионных моделей

Необходимые условия первого порядка:

$$\frac{\partial l(\beta,\sigma^2)}{\partial \beta'} = -\frac{1}{2\sigma^2}(-2X'Y + 2X'X\beta) = 0$$

$$\frac{\partial l(\beta,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(Y - X\beta)'(Y - X\beta) = 0$$

$$\Rightarrow \hat{\beta} = (X'X)^{-1}X'Y, \qquad \hat{\sigma}^2 = \frac{e'e}{n}, \qquad e = Y - X\hat{\beta}$$

Достаточные условия:

(см. Магнус и др., 2007, 10.5)

Общие свойства оценок МП

Если функция f(x) плотности распределения случайных величин X1,…Xn удовлетворяет некоторым условиям регулярности, то оценки МП обладают свойствами:

1) Инвариантность.

Если $\hat{\theta}_{МП}$ - оценка максимального правдоподобия параметра θ и $g(.)$ - непрерывная функция, то $g(\hat{\theta})$ является оценкой максимального правдоподобия параметра $g(\theta)$.

2) Состоятельность (определение состоятельной оценки дано в разделе 2.3.1).

3) Асимптотическая нормальность.

При $n \to \infty$ оценка вектора параметров имеет нормальное распределение

$$\hat{\theta}_{МП} \overset{a}{\sim} N(\theta, I^{-1}(\theta)),$$

Где $I(\theta)$ - информационная матрица Фишера, вычисляемая по формуле

$$I(\theta) = -E\left[\frac{\partial^2 l}{\partial\theta\partial\theta'}\right],$$ l — функция правдоподобия, $\frac{\partial^2 l}{\partial\theta\partial\theta'}$ - матрица Гессе.

4) Асимптотическая эффективность.

Оценка ММП дисперсии каждого параметра (один из диагональных элементов ковариационной матрицы $I^{-1}(\theta)$) является нижней границей для всех состоятельных асимптотически нормальных оценок этого параметра.

Проверка линейных гипотез с помощью теста Вальда

$$Y = X\beta + \varepsilon,\ \mathrm{var}[\varepsilon] = \sigma_\varepsilon^2 I$$

$$H_0 : Q\beta = q,\ ranqQ = r$$
$$H_1 : Q\beta \neq q$$

$$\varepsilon \sim N(0, \sigma_\varepsilon^2 I) \Rightarrow$$
$$\hat{\beta}_{ML} \sim N(\beta, \sigma_\varepsilon^2 (X'X)^{-1}) \Rightarrow$$

$$Q\hat{\beta}_{ML} \sim N(Q\beta, Q\,\mathrm{var}[\hat{\beta}_{ML}]Q')$$

$$W = (Q\hat{\beta} - q)'(Q\,\mathrm{var}[\hat{\beta}_{ML}]Q')^{-1}(Q\hat{\beta} - q) \sim \chi^2(r)$$

Проверка нелинейных гипотез с помощью теста Вальда

$$Y = X\beta + \varepsilon,\ \operatorname{var}[\varepsilon] = \sigma_\varepsilon^2 I$$

$H_0: g_j(\beta) = 0,\ j = 1, \dots, r$
$H_1: \exists j: g_j(\beta) \neq 0$

$$r = 1,\ g(\beta) \approx g(\hat\beta) + \frac{\partial g}{\partial \beta}(\hat\beta)(\beta - \hat\beta)$$

$$W = g'(\hat\beta)\left(\frac{\partial g}{\partial \beta}(\hat\beta)\operatorname{var}[\hat\beta_{ML}]\frac{\partial g'}{\partial \beta}(\hat\beta)\right)^{-1} g(\hat\beta) \sim \chi^2(r) = \chi^2(1)$$

Проверка нелинейных гипотез с помощью теста Вальда

$$Y = X\beta + \varepsilon,\ \operatorname{var}[\varepsilon] = \sigma_\varepsilon^2 I$$

$$H_0: g_j(\beta) = 0,\ j = 1, \dots, r$$
$$H_1: \exists j: g_j(\beta) \neq 0$$

$$r > 1,\ g(\beta) \approx g(\hat{\beta}) + \frac{\partial g}{\partial \beta}(\hat{\beta})(\beta - \hat{\beta})$$

$$W = g'(\hat{\beta})\left(\frac{\partial g}{\partial \beta}(\hat{\beta})\operatorname{var}[\hat{\beta}_{ML}]\frac{\partial g'}{\partial \beta}(\hat{\beta})\right)^{-1} g(\hat{\beta}) \sim \chi^2(r)$$

$$g(\beta) = \begin{pmatrix} g_1(\beta) \\ \vdots \\ g_r(\beta) \end{pmatrix}, \dots, \frac{\partial g}{\partial \beta} = \begin{pmatrix} \frac{\partial g_1}{\partial \beta_0} & \cdots & \frac{\partial g_r}{\partial \beta_k} \\ \vdots & & \vdots \\ \frac{\partial g_r}{\partial \beta_0} & \cdots & \frac{\partial g_r}{\partial \beta_k} \end{pmatrix}$$

Достоинства и недостатки теста Вальда

Достоинство: используются только оценки модели без ограничений

Недостаток: неинвариантность к способу параметризации

Проверка гипотез с помощью теста отношения правдоподобия

$Y = X\beta + \varepsilon,\ \mathrm{var}[\varepsilon] = \sigma_{\varepsilon}^{2} I$

$H_0: g_j(\beta) = 0,\ j = 1, \dots, r$
$H_1: \exists j: g_j(\beta) \neq 0$

$LR = -2(\ln L\,(\hat{\beta}_{\mathrm{R}}) - \ln L\,(\hat{\beta}_{\mathrm{UR}})) \sim \chi^2(r)$

В этом тесте используются оценки модели с ограничениями и без ограничений

Проверка нелинейных гипотез с помощью теста множителей Лагранжа

$Y = X\beta + \varepsilon,\ \mathrm{var}[\varepsilon] = \sigma_\varepsilon^2 I$

$$H_0: g(\beta) = \begin{pmatrix} g_1(\beta) \\ \vdots \\ g_r(\beta) \end{pmatrix} = 0,$$

$H_1: \exists j: g_j(\beta) \neq 0, j = 1, \ldots, r$

$H(\beta, \lambda) = l(\beta) - \lambda' g(\beta) \to \max$

$$\frac{\partial H}{\partial \beta_k} = \frac{\partial l}{\partial \beta_k} - \lambda' \frac{\partial g}{\partial \beta_k}$$

$$\frac{\partial l(\hat{\beta}_R)}{\partial \beta_k} - \lambda' \frac{\partial g(\hat{\beta}_R)}{\partial \beta_k} = 0$$

Если ограничения выполняются, то все множители Лагранжа равны 0, поэтому и градиент логарифмической функции правдоподобия близок к 0.

Проверка нелинейных гипотез с помощью теста множителей Лагранжа

$$Y = X\beta + \varepsilon,\ \text{var}[\varepsilon] = \sigma_\varepsilon^2 I$$

$$H_0: g(\beta) = \begin{pmatrix} g_1(\beta) \\ \vdots \\ g_r(\beta) \end{pmatrix} = 0,$$

$$H_1: \exists j: g_j(\beta) \neq 0, j = 1, \dots, r$$

$$LM = \left(\frac{\partial l}{\partial \beta}(\hat{\beta}_R)\right)' I^{-1}(\hat{\beta}_R)\left(\frac{\partial l}{\partial \beta}(\hat{\beta}_R)\right) \sim \chi^2(r)$$

В этом тесте используются только оценки модели с ограничениями

Расстояние Кульбака-Лейблера

f(x) – функция плотности истинного распределения, g(x|θ) – функция плотности оцениваемого распределения.

Расстояние между истинной и оцениваемой моделями Кульбака-Лейблера (Kullback and Leibler, 1951)

$$I(f,g) = \int f(x) \log\left(\frac{f(x)}{g(x|\theta)}\right) dx.$$

Расстояние Кульбака-Лейблера

$$I(f, g) = \int f(x) \log\left(\frac{f(x)}{g(x|\theta)}\right) dx.$$

$$I(f, g) = \int f(x) \log(f(x)) dx - \int f(x) \log(g(x|\theta)) dx$$

$$I(f, g) = E_f[\log(f(x))] - E_f[\log(g(x|\theta))],$$

Расстояние Кульбака-Лейблера

$$I(f,g) = \int f(x)\log(f(x))dx - \int f(x)\log(g(x|\theta))dx$$

$$I(f,g) = E_f[\log(f(x))] - E_f[\log(g(x|\theta))],$$

$$I(f,g) = C - E_f[\log(g(x|\theta))],$$

$$C = \int f(x)\log(f(x))dx$$

Критерий Акаике выбора модели

$$I(f, g) = C - E_f[\log(g(x|\theta))],$$

$$C = \int f(x)\log(f(x))dx$$

Критерий Акаике основан оценке

$$E_y E_x[\log(g(x|\hat{\theta}(y)))],$$

где y – это данные.

Критерий Акаике выбора модели

Критерий Акаике основан на оценке

$$E_y E_x[\log(g(x|\hat{\theta}(y)))],$$

где y – это данные.

Основной результат Акаике:

$$\log(\mathcal{L}(\hat{\theta}|\text{data})) - K = C - \hat{E}_{\hat{\theta}}[I(f, \hat{g})]$$

Критерий Акаике выбора модели

Основной результат Акаике:

$$\log(\mathcal{L}(\hat{\theta}|\text{data})) - K = C - \hat{E}_{\hat{\theta}}[I(f, \hat{g})]$$

$$\text{relative } \hat{E}(K\text{-}L) = \log(\mathcal{L}(\hat{\theta}|\text{data})) - K.$$

$$\text{AIC} = -2\log(\mathcal{L}(\hat{\theta}|\text{data})) + 2K.$$

Чем меньше AIC, тем лучше модель.

Для линейной регрессионной модели с нормально распределенными ошибками:

$$\text{AIC} = n\log(\hat{\sigma}^2) + 2K.$$

Критерий Байеса (Шварца) выбора модели

Schwarz (1978) derived the Bayesian information criterion as

$$\mathrm{BIC} = -2\ln(\mathcal{L}) + K\log(n).$$

Чем меньше BIC, тем лучше модель.

Thank you for your attention!

20, Myasnitskaya str., Moscow, Russia, 101000
Tel.: +7 (495) 628-8829, Fax: +7 (495) 628-7931
www.hse.ru